


Looking across the States: Perspectives on School Accountability

By Casey D. Cobb

First-generation models of educational accountability were mainly bureaucratic and regulatory in nature. Most U.S. states required public school districts to comply with various rules and process standards, such as a fixed number of school days per year and minimum pupil-to-teacher ratios. On a national scale, the accountability movement of the 1970s and 1980s evidenced a shift from an emphasis on rules to a focus on results (Elmore, Abelman, & Fuhrman, 1996). During this period, large-scale minimum competency tests served as the primary mechanism for school and district accountability. More recently, new systems of educational accountability have evolved—including more comprehensive performance-based models.

The emergence of state-level, performance-based accountability systems is a predictable consequence of the standards and assessment movements in education. By the end of the last decade, 49 U.S. states had developed a set of learning standards (“Seeking stability for standards-based education,” 2001)—what stu-

 Casey D. Cobb is an assistant professor of educational policy in the Department of Educational Leadership of the Neag School of Education at the University of Connecticut, Storrs, Connecticut.

dents were to know and be able to do at specific grade levels—and most states had attempted to align their assessment systems with those standards. With the learning objectives and measures of performance in place for students, state governments had the means to hold schools accountable for student performance. Indeed, prior to 2001 performance-based accountability policies were operating in 33 states (Goertz & Duffy, 2001). With the advent of the No Child Left Behind Act of 2001 (NCLB), *every* state was required to adopt an accountability system. The federal legislation even provided a prescriptive set of guidelines for states to follow.

The combination of an onerous set of federal requirements and a stressful fiscal climate prompted many states to acquiesce to the NCLB framework as their new accountability policy. A small number of state accountability policies (Texas, Delaware, and Florida among them) did not change much, as they were already aligned with key aspects of NCLB. Other states (e.g., Colorado, Arizona) opted to maintain dual systems of accountability: one that satisfied the specific requirements of NCLB and another that preserved the state's existing approach to accountability.

This article presents a framework to evaluate the policy choices sanctioned by state systems of performance-based accountability. Key components and mechanisms of both the NCLB Act and various state accountability policies illuminate the fundamental means (and differences) by which schools are evaluated and held accountable. The paper concludes with a discussion of the policy options available to policy makers of such systems and puts forth a set of recommendations most likely to effect improvement in schools. First, I begin by defining performance-based accountability and describing the basic elements of state systems of accountability that are in practice today.

Background

Conceptions of Accountability

Although misperceived as such, accountability is not a monolithic construct. On the contrary, accountability has been defined in a variety of ways. Elmore et al. (1990) tracked the evolution of accountability policies, highlighting three disparate theories of accountability: technical-based accountability, client-based accountability, and professional-based accountability. The technical approach assumes that improvement will take place only if teaching and learning practices are grounded in scientifically-based knowledge. The cyclic process of collection, analysis, and reporting of educational performance indicators that are aligned with clearly defined performance goals constitute a viable accountability system. In contrast, the client perspective holds that schools will improve their performance when educators hold themselves accountable directly to their clients (i.e., students, parents, and the community). Lastly, professional-based accountability occurs when school practitioners and leaders are afforded opportunities to make decisions,

develop expertise, and maintain autonomy in their field of work. Formal policies at the state and federal levels suggest we are amidst an era of technical-based accountability. Arguably, this current emphasis diminishes professional and client-oriented forms of accountability. Other theorists have offered complementary views of accountability (see, e.g., Gintis, 1995; Glass, 1972; Kirst, 1990; Levin, 1974; Macpherson, 1996; Newmann, King, & Rigdon, 1997).

Contemporary views of performance-based accountability systems suggest the presence of at least four components: (1) goals for student learning at all grade levels, (2) accurate measurement of student learning outcomes, (3) rewards for local educators for good student outcomes, and (4) interventions targeting failing schools (Scafidi, Freeman, & DeJarnett, 2001; Newmann, King, & Rigdon, 1997). Kuchapski (1998) broadened the basic elements of accountability to include the aspects of planning, reporting, monitoring, assessment, communication, and responsiveness. Macpherson (1996) conceptualized accountability as both criteria and process. Accountability *criteria* are the basis upon which decisions are made about effectiveness. Accountability *processes* refer to the manner in which data are collected, stored, analyzed, and reported as means to improve performance.

Accountability Systems in the U.S. States

The NCLB legislation has indeed cast considerable influence on state assessment and accountability programs. The law provides a fairly prescriptive set of guidelines for state accountability systems. However, both beyond—and to a lesser degree within—these NCLB requirements, states are still left with the constitutional authority to assess what, when, and how they want, and to decide what to do with this information. That said, realistically, state options are limited by available resources, overall commitment to education, and other capacities—factors that vary from state to state.

NCLB notwithstanding, state-level accountability systems appear to subscribe to the same general formula. Academic learning standards are established in certain content areas, state assessments are developed or revised to align with those standards, and criteria are used for judging and establishing consequences for performance. However, the process is not as straightforward as it may seem.¹ Nor is it consistently implemented across the states that have such systems. For instance, not all state proficiency standards are assessed by state tests. Rhode Island maintains curriculum frameworks in mathematics, science, English language arts, social studies, arts, family/consumer science, and health education, but administers formal assessments in only three of those areas (mathematics, English language arts, and health) in addition to a writing assessment. Similarly, Vermont's "Framework of Standards and Learning Opportunities" is more extensive and far-reaching than what is assessed by way of its state assessment system. (For a comprehensive inventory of state assessment and accountability systems, see Goertz and Duffy (2001).)

The focus of most state accountability systems is on the performance of schools or districts. Several state policies also dictate important decisions regarding students such as grade level promotion and awarding of diplomas. Twenty-three states have or will soon have graduation exit exams (Clarke, 2001). A small group of states holds teachers responsible for student performance. For example, Georgia, Kentucky, and North Carolina provide rewards for educators based on student exam results.

“School performance” is typically represented by the aggregation of student indicators, such as average test scores, percentages of students scoring at proficient level, and attendance rates. It is assumed the school is collectively and directly responsible for these performance measures (or student outcomes, as the case may be).

Student performance can be measured in many ways. State curriculum standards tend to reflect at least some of this variety. For instance, several states establish goals in the cognitive, affective, behavioral, and social domains.² Not all these areas are readily measurable, especially by way of large-scale state assessments. Locally-based assessments may be used to supplement or in lieu of state administered assessments.

Standardized assessments are by far the most common instrument states use to measure student learning and evaluate schools. Forty-eight states use a state assessment as the “principal indicator of school performance” (Goertz & Duffy, 2001, p. 2). Eleven states use test results *exclusively* to rate their schools (Quality Counts, 2001, p. 9). Indicators of student performance typically fall under the academic domain, although some states are also interested in tracking non-cognitive student outcomes and other indicators of institutional quality (e.g., percentage of students pursuing post-secondary education, dropout rate). State-administered assessments tend to be viewed as uniform from state to state, although they can vary in several respects (e.g., reference base of the test, types of test items, and the grade levels and subject areas assessed).

A Clarifying Framework

In spite of the heavily prescriptive NCLB Act, states remain responsive to their own unique cultural norms and traditions in their efforts to evaluate school performance. State accountability policies reflect these influences. Fundamental questions arise around schemes for holding schools accountable. What counts as school performance and how is it assessed? Is the goal to maximize performance or minimize failure? At its core, what is the purpose of the accountability system? What happens when schools fail to meet expectations—and who is ultimately in charge of all this?

This section articulates the type and range of policy choices available to state policymakers. Specifically, I present a framework to expand theoretical understanding of performance-based accountability.³ Six key dimensions underlie large-scale policies on educational accountability: *definition of performance, assessment of*

performance, goal orientation, evaluative function, consequential nature, and locus of control. These criteria can be used to clarify and compare performance-based accountability systems. A brief description of the framework follows, along with examples of specific state and federal accountability policies that help shed light on each dimension.

Definition of Performance—What is to be assessed?

The first question is *what* counts as school performance? What is worthy of being held accountable? Accountability policies vary along this fundamental dimension. States deem what knowledge, skills, and proficiencies are most important by way of their accountability systems. On one end of the spectrum are broad, global conceptions of performance. On the other are narrow, specific conceptions (see Table 1 for a summary of the range of characteristics for each dimension). NCLB requires schools to assess student performance in mathematics and the language arts, and later will include science. NCLB also asks schools to track graduation rates and allows elementary and middle level schools to choose an additional indicator of school performance.⁴ Generally speaking, Delaware, Vermont, North Carolina, and Texas emphasize the fundamentals of reading, writing, and mathematics. In addition to these core academic skills, Kentucky, Nebraska, and Iowa assess performance in other traditional subject areas (e.g., science, social studies). Missouri and Rhode Island treat student performance more holistically, extending their

Table 1
Conceptual Framework for Performance-Based Accountability Systems
in Education

Accountability Dimension	Policy Choices
Definition of Performance	narrow broad global particular
Assessment of Performance	multiple unitary measures holistic academic standardized authentic
Goal Orientation	excellence equity minimum competency maximum potential
Evaluative Function	oversight improvement summative formative
Consequential Nature	punitive supportive high stakes low stakes
Locus of Control	external internal centralized decentralized

definitions beyond the academic domain and assessing such areas as health and physical education. Other nontraditional school performance indicators that are broadly conceived include measures of school climate and parental involvement. Fiscal constraints and the sheer volume of the operation restrict and place limits on what is measured and, thus, what ultimately represents performance.

Another often overlooked question is whether schools themselves are assessed or whether student assessments are used as a proxy for school performance. The line between school and student performance is blurred. Most school indicators are derived from pooled student data. School performance in mathematics is represented by some aggregate measure of student performance on a math exam (e.g., average score, percentage scoring at proficient level). There are, of course, exceptions (e.g., school climate, levels of parent involvement indicators).

Assessment of Performance—How is performance assessed?

Not only what is assessed is important, so too is *how* it is assessed. We all have a reasonable notion as to what “reading achievement” represents; however, the true source of its meaning lies in its method of measurement. How is the construct operationalized? How are students required to demonstrate their ability to decipher text and interpret its meaning? What measure or measures are used?

Assessment techniques may vary. Are multiple instruments used or is there a single measure of performance? Does a state rely on standardized, group assessments that place test-takers in hypothetical situations, or does it also invest in more authentic measures of student performance? Is assessment continuous and ongoing, or is it conducted more periodically?

A range of assessment instruments are available. Examples include standardized exams, student portfolios, student work samples, performance observations, and writing prompts. Most states rely primarily or exclusively on standardized exams. Although standardized tests share many commonalities (e.g., on-demand, timed, paper-and-pencil), there can be some deviation in exam format and item type. As an example, Missouri’s academic subject exams employ three item types—multiple choice, constructed-response, and a performance event—each of which takes about an hour to complete. In the performance event section, students must show their work or explain how they arrived at their answers.

Standardized measures of performance represent a unitary approach to assessing student achievement. North Carolina and Texas, and to some extent Delaware, adhere to this strategy. In modest contrast, Vermont tests in the core basics of reading, writing, mathematics, and science, but permits the use of other assessment instruments, such as locally developed criterion-referenced exams, commercially developed norm-referenced exams, and student portfolios. Maine has invested heavily in a local assessment program (Maine Department of Education, 2004).

NCLB asks states to administer standards-based assessments in major subject

areas. In addition, NCLB requires all states to participate in the National Assessment of Educational Progress 4th and 8th grade reading and math tests. To satisfy these requirements, few states will stray far from standardized test batteries. Economic realities will continue to compel states to opt for cost-conscious assessment systems, such as once-a-year, machine-scored standardized tests. Witness New Hampshire, Rhode Island, and Vermont, which have formed a partnership to develop common assessments for grade levels that were not previously part of their testing programs (New England Compact, 2004).

Goal Orientation—Where’s the bar (and how many bars are there)?

Two core American values—equity and excellence—compete for prominence in the arena of educational accountability. *Equity* was the focus of educational policy in the 1960s and 70s. The pursuit of educational *excellence* surfaced amid (and in response to) accusations of a “rising tide of mediocrity” (A Nation at Risk, 1983). Appeals to excellence are apparent among the rhetoric of reform efforts of today. Many policies strive to simultaneously pursue both equity and excellence, despite the illogic of their complementarity (Fritzberg, 2000).

Is the goal to achieve minimum proficiency for some, or to demonstrate continuous progress for all? Are the lowest performers—whether they be students or schools—the *real* focus of the accountability policy? Are medium and high performers equally a part of the accountability policy? Most state accountability systems uphold a minimum performance threshold (e.g., “all students in every school must perform at the proficient level in reading by 2013-14”). That said, *proficiency* can mean something different to each state. The number of students (or schools) truly affected by such proficiency standards depends on where the bar is set.

State policies tend to fall toward the “minimum-competency” end of the spectrum. Missouri’s overall objective is for students to achieve at least at the proficient level. Missouri, however, also evaluates progress by inspecting change between high and low scoring students. Progress is indicated by fewer students scoring in the two lowest performance categories and by more students falling in the two top levels. Vermont recognizes students that achieve standards “with honors.” Similarly, Delaware recognizes students who score in the highest performance category (Level 5), rewarding them with “Distinguished Performance Certificates” and, in some cases, \$1,000 scholarships. Delaware also awards three different types of high school diplomas, depending on scores on the statewide test.

NCLB legislates that all students reach “proficiency” by 2013-14, but leaves it to individual states to determine performance benchmarks for proficiency. We have witnessed great variation in states’ meaning of proficiency. Higher performance levels exist above proficiency (e.g., “exceeds standards”) suggesting that NCLB leans more toward a minimum (or at least *non*-maximum) goal orientation.

Evaluative Function—What is the purpose?

The evaluative function criterion speaks to the underlying purpose of the accountability system. Is it designed to improve schools or to monitor them? One way to think about this is the degree to which an accountability system embraces formative and summative forms of evaluation. With a formative approach, the evaluation is intended as the basis for improvement. Summative evaluation, as the name implies, is about drawing final conclusions about performance. The two concepts are not necessarily mutually exclusive; both can be pursued in any evaluation. Indeed, formative evaluation “is, to a large extent, best designed as summative evaluation of an early version, with particular attention to components or dimensions rather than a holistic account (because this facilitates improvement)” (Scriven, 1997, p. 498).

On their face, state accountability strategies would appear to be more summative in nature. Several state policies, however, adopt a formative philosophy. Missouri requires schools to develop comprehensive improvement plans. Iowa asks something similar of its school districts. Both states appear to make it a priority not only to judge institutional performance, but to use the information to improve practice. The efficacy of such school improvement plans remains unsubstantiated and probably varies from school to school, plan to plan.

While its symbolic intent may be rapid school improvement, in reality NCLB is predominantly a monitoring policy (i.e., summative). There is no guidance—and no mechanism—to help schools close achievement gaps, improve instruction, or make schools safer. The optimist would find value in the flexibility and autonomy afforded to states and schools. The cynic would refer to this as an unfunded and unsupported mandate. Whichever the position, NCLB is decidedly more summative than formative.

Consequential Nature—What happens?

Consequences are a key aspect of performance-based accountability systems; consequences give accountability, in its crudest form, “teeth.” Something must happen in the event that schools do not achieve to that which they are obligated. Policy levers come in the form of punishments (e.g., probationary status, school audits), in the form of inducements (e.g., school monetary awards, student scholarships), and in the form of support (e.g., additional resources, professional development collaboratives). Many are of the high-stakes variety for students (e.g., grade promotion, high school exit exams) and schools (e.g., state takeover). At the other end of the continuum are “less high-stakes” consequences, such as the public reporting of school test results.

By 2008, at least 28 states will have in place a high school graduation test (Goertz & Duffy, 2003). Delaware uses a single indicator of performance—standardized exams in basic subject areas—to make high stakes decisions about students

and schools. For instance, students in grades 3, 5, and 8 whose Delaware state test scores fall below standard (i.e., Level I) in reading are required to attend summer school (Delaware Department of Education, 2004). Similarly, North Carolina and Texas use test scores to make decisions regarding student grade promotion and graduation.

Provisions in NCLB call for serious consequences for schools that fail to meet adequate yearly progress (AYP) for two consecutive years. Those consequences include offering specialized tutoring services and intradistrict school choice, and may eventually involve the complete reconstitution of the school.

Locus of Control—Who's in charge?

By their very nature, state-level assessment and accountability programs are centralized forms of control. *State* standards are measured via *state* assessments. However, a number of states—Iowa, Nebraska, Vermont, and Rhode Island among them—cede more authority to local education agencies. For instance, Nebraska and Iowa required tests at certain grade levels or grade spans, but left it to school districts to design the assessments (Goertz & Duffy, 2003). Maine has similarly endorsed locally developed instruments. Most states, however, exhibit strong forms of external control over the assessment and accountability of schools and students.

Nebraska's STARS (School-based Teacher-led Assessment and Reporting System) program uses a balance of classroom and standardized assessments to inform and motivate. STARS intends to foster capacity building and strives to enhance "assessment literacy" among educators in that state. Missouri trains their teachers to learn about, develop, use and score performance-based assessments; it is part of professional development and part of an assessment-minded culture. Missouri relies heavily on school districts to devise their own assessment plans to measure progress toward state standards that cannot be (or are not) tested via statewide exams. Rhode Island's SALT (School Accountability for Learning and Teaching) program is referred to as "a school-centered cycle of activities to improve school and student performance."⁵ SALT requires districts to engage in various self-study activities aimed at school reform. Rhode Island also asks districts to set their own performance targets. Vermont has its own state assessment, but offers the option of locally-derived assessments. Delaware, Texas, North Carolina, and Kentucky take a very centralized approach, which is consistent with the prescriptive nature of NCLB. NCLB affords little discretionary authority over the assessment and evaluation protocols of schools receiving Title I monies.

Policy Options within the Framework

Education policies are presumably written to achieve specific goals. Certain policy mechanisms are more effective than others at reaching their objective. What are justifiable positions within each of the six dimensions? An accountability

system could define performance *narrowly*, measure it *holistically*, and use the results *formatively* to shape school reform—all the while exhibiting a strong form of *centralized* control.

Attempts should be made to carefully consider the consequences of certain policy choices. For instance, the combination of holding schools accountable to the basics of reading, writing, and mathematics and attaching high-stake consequences to performance runs the risk of squeezing out other important student outcomes (cf. McNeil, 2000). On the other hand, holding schools accountable to a focused set of goals that are central to its mission—and attaching strong incentives and disincentives—may serve as an effective policy lever for reform.

The framework presented here permits judgments to be made about the choices within each dimension. For instance, one might argue that formative approaches are more likely to effect real change than strategies that rely exclusively on pressure and the threat of punitive consequences. As Scriven (1997) observed, “The role of formative evaluation is to provide feedback on midstream merit, as a service to assist program improvement” (p. 499). What follows is a discussion of the various policy choices available to policy makers of accountability systems and their probable consequences. I put forth a set of policy recommendations that are most likely to bring about school improvement.

Definition of Performance

There are tradeoffs to choosing either a broad or narrow definition of performance. A focus on core academic skills exemplifies a narrow definition.⁶ One advantage of a focus on academic skills is that there is little disagreement over the desire for students to acquire such “basic skills.” Every child should know how to read and write, add and subtract. Other advantages to limiting the scope of performance assessment are economic—there is simply less to measure. All other things equal, the fewer the areas of performance, the less burden on already over-taxed state agencies of education that, in addition to measuring performance, are also asked to provide technical support to those schools not making the grade.

The disadvantage of narrow conceptions of performance is, of course, that they can be overly narrow. In terms of students, there are myriad skills, behaviors, and habits of mind that are of value—many of which fall outside the cognitive domain.⁷ In terms of schools, there are many important functions they serve for students beyond developing academic skills. If such outcomes truly do matter, then these too should be part of a comprehensive accountability system. Focusing on one aspect of schooling can serve to de-value other important purposes of education.

It is ingrained in the educational establishment that performance be defined by discipline or subject area (e.g., math, history). We first determine what students should know within a domain of knowledge, and what skills they should be able to demonstrate. If, instead, we asked what students should know and be able to do

irrespective of academic subject areas, we might articulate outcomes such as “make a persuasive argument or speech,” “be a creative problem-solver,” “demonstrate higher-order thinking skills.” While such laudable objectives fill school mission statements, they are given little if any credence in present day accountability systems. Indicators of performance could look much different.

Take, for instance, NCLB, which asks students to pass state tests in reading/language arts and mathematics—core academic basics. The problem with such an approach is that it rests on the fallacy that the basics are the “building blocks” to more complex ways of thinking. It is an erroneous belief of educational practice to delay experiences that foster expert thinking until after the “basics” have been learned (Brooks & Brooks, 1993; Gardner, 1991; Haas et al., 2004). Expert abilities involve being able to actively question and explore, and for an individual to understand “a concept, skill, theory, or domain of knowledge to the extent that he or she can apply it appropriately in a new situation” (Gardner, p. 119). Expert thinkers have “the ability to think and act flexibly with what one knows” (Wiske, 1997, p. 40). Requiring students to first acquire basic skills, void of an authentic context with which to apply them, severely limits their opportunities to practice and develop expert ways of thinking. For a school to set its sights on the narrow outcomes of basic skills, such an approach runs contrary to good educational practice. We may see short term improvement in test performances, but these advancements represent false markers of genuine student learning.

Assessment of Performance

The notion of “school performance” has reached reified status. We take for granted that a school deemed “high-performing” is, indeed, just that. The same goes for “low-performing” or “failure” schools that qualify for such distinctions through sophisticated scoring mechanics. One may question, however, whether “school performance” is something measurably observable, or whether it is, like so many other social science constructs, an undocumentable abstraction.

Within accountability systems, assessments of school performance almost always entail aggregate measures of student scores on standardized tests. Assessments must adhere to particular standards of reliability and validity. For this and for economic reasons, most all states administer standardized test batteries comprised of multiple choice, short answer, and to a lesser extent written response items. Single measures of assessment, however, suffer from the same problems of narrow definitions of performance noted above.

Accountability systems are in large part defined by methods of assessment. State-administered assessments can provide important diagnostic information to schools, but can end up forcing upon schools an overly narrow method of assessing student and school progress. Accountability systems, especially those that attach high stakes to performance, should be careful of relying too heavily on single, large-

scale assessments as a mechanism for school reform. State tests might be better used diagnostically at both the group and individual levels. Although there are obvious economic tradeoffs, the use of multiple measures of assessment greatly improves test validity.

In the final analysis, one must consider the *validity* of the inferences drawn from scores on these measures of performance. Validity refers to the meaningfulness, usefulness, and appropriateness of those inferences. Student scores on state assessments may not be appropriate indicators of school—or even student—performance. In order to be valid measures of school performance, there must be some certainty that student performance is directly linked to what schools do for, to, or with students. To be valid, scores on annual, on-demand, three-hour paper-and-pencil exams should represent a meaningful, useful, and appropriate measure of student achievement. It is highly questionable whether standardized test scores by themselves represent valid indicators of school performance, let alone measure authentic student learning. Singular measures of student and school performance may be fiscally responsible, but the educational costs associated with such narrow assessment strategies make them educationally reckless. If multiple measures cannot be used in an accountability system, then perhaps no single measure should be used at all.

Goal Orientation

States with rigorous performance standards tend, unsurprisingly, to have high proportions of students achieving below the proficient level. Such scenarios are not well-received politically and can potentially result in less demanding standards, easier statewide tests, or delayed performance expectations (e.g., prior to NCLB, Kentucky expected each school to reach the proficient level by 2014).⁸ Although “minimum standards” does not necessarily mean “low standards,” aggressive minimum proficiency targets tend to soften to the point where politically and economically acceptable levels of students (or schools) can meet them. As but one example, Massachusetts recently established an appeals process to allow some students to graduate even if they failed the 10th grade state graduation test.⁹ Thus, under such policies, only a portion of students (or schools) is affected—though arguably those most in need of improvement.

Few state policies seek to maximize the performance of all parties, no matter where they fall on the performance spectrum. Pursuing both the minimum and maximum goal orientations simultaneously can be problematic. To the extent that *equity* and *excellence* represent conflicting pursuits, one invariably takes prominence over the other. For instance, states that are interested in reducing the achievement gap between high and low poverty students may find it difficult if parties on the upper end continue to excel.

A “minimum proficiency” goal orientation has much to offer. The best case scenario is that the external pressure will lead to improved performance (this, after

all, is the rationale behind such systems.) Clear and measurable benchmarks are established for students and schools. On the other hand, focusing limited resources on the performance of the lowest achievers can take attention away from maintaining the higher levels of performance of those who have already met the minimum threshold. A minimum proficiency orientation can also result in unproductive behaviors on the part of over-pressed students, teachers, and administrators.

A “maximum for all” approach has the advantage that each student or school is part of the accountability process. Each is held accountable, at least in theory, to its potential. But it is difficult to measure potential, especially by way of traditional assessments. The comparison between the minimum and maximum orientation can be likened to the now dated comparison between American and Japanese business models. Under the American approach—management by financial objectives—specific budgetary goals (e.g., profit margins) would drive business operations. In contrast, Edwards Deming’s Total Quality Management (TQM) approach focused more on the process of building a good product, and assumed that financial success would result.

As an accountability system, NCLB really takes a minimalist approach. Every child must achieve “proficiency” on state standardized assessments. However challenging a goal for all children to meet, proficiency is still a minimum level of competency. Again, not necessarily low, but minimum. The unfortunate illusion is that even if all children were to attain this level of proficiency, it does not necessarily lead to improved or more equitable life outcomes for those children. Minimum goal orientations also tend to take a one-size-fits all approach. The point here is less about “minimum” levels of proficiency and more about a “standardized” level of performance. Requiring every student to meet a common outcome compromises any role of the school to develop in kids their own unique talents and abilities. A standardized approach devalues the individuality of the student. Alternatively, a non-minimum, non-standardized goal orientation permits a curriculum and assessment system that is more conducive to a diversity of student interests, talents, and abilities. Multiple ways of being smart are rewarded under approaches that seek to maximize student potential, however construed.

Evaluative Function

Perhaps the primary role of an accountability system is to serve as monitor, and to ensure compliance to established performance standards. Meeting a set of externally defined standards requires the oversight of an independent, external entity. At least that is how most performance-based accountability systems are designed. If they do their job, they provide an important external check on the performance of publicly-supported schools.

It is also arguable that accountability systems have the intention of improving the performance of schools. Some, however, are less structured and less capable of

achieving this ideal. Accountability policies may place more emphasis on monitoring than, say, capacity building.

Macpherson's (1996) views accountability as a process "to collect, store, report, and use data to improve the quality of performance and services" (p. 20). In this sense, the accountability agency serves as informant as opposed to overseer. States take on the role of providing assessment information to schools. The assessments provide diagnostic information and are designed to enlighten classroom educators, building and district administrators, and state education officials as to strengths and areas needing improvement. Several states provide or facilitate assistance through regional professional development centers. These professional networks are typically staffed by experienced teachers and state department personnel and they primarily serve under-performing schools.

A strictly or even predominantly summative approach to evaluating schools may not necessarily lead to improvement in performance. As Don Graves has astutely claimed, testing isn't teaching. It is useful for a school to know that a large proportion of its third graders scored poorly on a state language arts exam. The assessment identifies a potential problem area in that school. The school may be prompted to take action, but the test data themselves do not provide the blueprint nor do they reveal factors explaining the low scores. Let's say that one of the root causes has to do with low student attendance. If the school focuses its effort on refining its curriculum and instruction—after all, the students are not performing well here—it is missing a major source of the problem. A lack of organizational capacity could also contribute to the low student scores. Or inadequate teacher skills. It could be that the teachers lack sufficient training in teaching to the state curriculum standards.¹⁰ Other potential contributors to poor test scores include low student engagement, low community expectations, inadequate levels of parental involvement, and poor school climate.

The point here is not to blame the messenger, but rather to say that the messenger as currently conceived is not necessarily equipped to solve the problems of schools and may in fact distract schools from seeing or exploring the multifaceted complexity of problems that low performing schools invariably face. Summative test scores tend to focus schools' efforts on reversing the status in this domain of performance and to seek short term fixes to the problem. Formative assessments, by their very nature and through the information they offer, permit schools to act as problem-solvers as opposed to myopic reactors.

Consequential Nature

Accountability systems are designed, in part, to prompt schools to perform at desired levels. Changing school behavior and the people that work in them is challenging business to say the least. Inducements and punishments attached to performance standards are part and parcel of today's performance-based systems.

This represents a behaviorist approach to organizational reform. The thread of stakes attached to organizational performance indicators is a form of external pressure that is assumed to spur changes in behavior. Should states be in the business of punishing or rewarding students and schools for performance? And more importantly, is this an effective process? Do incentives and disincentives positively alter the behaviors of educators? Do the changes translate into increased student learning? Do the indicators or the measures themselves lead to consequences that are desirable or undesirable? For instance, among the unintended consequences of high stakes testing arrangements have been an over emphasis on rote memorization and the narrowing of the curriculum to only those subjects tested (e.g., McNeil, 2000; Taylor, Shepard, Kinner, & Rosenthal, 2001).

In a national poll of teachers, almost half reported spending “a great deal” of time on test taking skills.¹¹ Nearly two-thirds of the teachers indicated that the state tests had forced them to concentrate on the material tested. For some, “teaching to the test” can be construed as a positive consequence, but the idea loses value when the test becomes so emphasized as to marginalize untested aspects of the curriculum such as health practices, artistic and musical skills, and foreign language ability as well as other worthy student outcomes such as educational persistence, high aspirations, curiosity, cooperativeness, and citizenship.

Delaware uses a single indicator of performance—standardized exams in basic subject areas—to make high stakes decisions about students and schools. In contrast, under Colorado’s Basic Literacy Act, school districts must use evidence from individual reading assessments in addition to results from the state assessment to make evaluative (proficiency) judgments of third graders.¹²

What of the consequences of the low-stakes variety? Public reporting of test scores falls in this category and has been going on for quite some time. Newspapers publish annual results of state exams. There are stakes due to the public nature of the data, and the fact that scores are presented side by side. Low scores can be a public embarrassment. The scores are viewed as proxies for school quality, and they are the only proxy to go by. This is low stakes relatively speaking because there are no explicit consequences attached to performance. There are no explicit goal targets to be met, nor explicit consequences for meeting or not meeting them. Complete reliance on public reporting of student performance to hold schools accountable may not provide enough incentive for every school to improve—at least not as public reporting is currently practiced. The notion of public reporting may require more teeth. For instance, full disclosure is one idea. Schools should not be permitted to selectively report on their performance. In other words, they should be required to disclose what is going well and what is not going so well.

Considering that state agencies are quite distanced from the everyday action and practice in schools, perhaps they are not in the best position to evaluate schools. It is a formidable task for state departments of education to accurately and validly assess and track the progress of thousands of students or hundreds of schools.

Adding strict consequences for poor performance applies pressures that could very well lead to undesirable consequences and ineffective practices. Snapshot, timed exams that sample domains of knowledge and skills should not be used for high-stakes decisions.¹³

State assessment data can be useful to schools, but the pressures to perform well on singular measures of performance has forced schools to focus more on the test results at the expense of good teaching and learning. There are countless anecdotal stories of teachers and administrators spending valuable time and resources on improving test scores. Under pressure, practitioners and those that lead them know how to bring about immediate improvements in test scores; devoting time during the school day to help students to *take* the tests will indeed raise scores, albeit marginally and artificially. Test scores become an end in and of themselves and not used as a measuring stick to validate what is happening in schools nor as a benchmark to document improvement. This again runs contrary to good education.

Locus of Control

Is the intent of accountability policies to keep tabs on how schools perform? Is it to evaluate student performance against an external set of standards (e.g., state curriculum)? To what extent do those being held accountable have control over the standards, assessments, and consequences for performance?

Formal, centralized systems of accountability dominate today's education policy landscape, where schools have little say over what debt to which they are held accountable. In externally-based systems of accountability, schools are in the position of being *held* accountable (see, e.g., Rallis & MacMullen, 2000). Prior to these systems, schools operated under a less formal, decentralized form of accountability. Under such conditions schools had the opportunity of *being* accountable. Although some schools acted more internally accountable than others, all schools at least had the opportunity to take ownership of educational accountability: what they were accountable for and to whom they were accountable.

Ablemann et al. (1999) reported that teachers feel a sense of responsibility for their students, and feel most accountable to their students and families—safety issues, socialization, caring for their well being of their students are paramount to scoring well on state exams. This is a very different conception of accountability than conveyed by large-scale policies on accountability. Such systems are characterized by a strong external component (e.g., *state*-applied pressure to perform to *state* expectations). While we have learned that states can and do differ in terms of their conception of performance, most value standardized test performance, and as such, view school accountability based on this factor.

If states are interested in comprehensively evaluating school quality, existing accountability models are probably not up to the task. Although entrusting schools and districts to be accountable to themselves may be too *laissez-faire* for some, states

might consider providing districts with a framework to evaluate themselves. Others have argued for local flexibility within comprehensive accountability systems.¹⁴ Policies in Nebraska, Iowa, and, to a lesser extent, Missouri, Vermont, and Rhode Island, appear to provide considerable autonomy to schools and districts.

The literature on school reform indicates that standards of success that are established externally are incompatible with lasting school change. Recent studies have indicated that external pressures alone may lead to changes in teaching of content but not of teaching pedagogy.¹⁵ It is not enough to give schools the rules of the game and then expect them to take action in the absence of adequate resources. One must not ignore solid research on what leads to real reform in schooling communities. Real and lasting change emerges from within.¹⁶ School reform models such as the Coalition of Essential Schools, Accelerated Schools, and Comer's School Development Program have demonstrated the value of a grassroots approach to reform. Along these lines, Vermont's guide to action planning acknowledges that "action plans developed at the school building level have the greatest impact on student achievement when those closest to the implementation of strategies, teachers, are members of the action planning team."¹⁷

The Annenberg Institute asserts that the fundamental work of accountability is the "continuous and reflective use of data."¹⁸ Annual, three-hour, answer-on-demand tests are not consistent with this definition. Rather, *continuous* means assessment and accountability is embedded in classroom, school and district practice. *Reflective* means those at the street-level reflect and act on the assessment information. State assessment reports that refer to the percentages of students that are deficient in reading or that are not performing at grade level are only of limited use to schools in need of significant change. This is not the kind of information that can be reflected upon or used in a continuous fashion.

Finally, it is important to note that the NCLB Act has cast a dark shadow on the future of state and local control over education. Federal mandates of annual testing in designated subject areas with specific reporting requirements has forced the hand of state education agencies. Clearly, less control is available at the state and local levels.

Conclusion

Six important dimensions underlie contemporary systems of accountability: definition of performance, assessment of performance, goal orientation, evaluative function, consequential nature, and locus of control. Each dimension constitutes a range of choices.

Performance-based accountability policies can be clarified, compared and evaluated along these six criteria. Are schools held accountable by way of a supportive or punitive approach? Via a local policy or state mandate? Equally important are the criteria on which schools or students are evaluated. What skills,

behaviors, dispositions, and knowledge are considered important and how do states choose to measure these outcomes? Is performance broadly or narrowly defined? Broadly or narrowly assessed? Performance-based accountability systems also differ in terms of their intent and underlying philosophy. For instance, accountability policies may vary in the degree to which they embrace accountability for accountability sake (i.e., purely a monitoring function) and the degree to which they pursue school improvement.

The framework opens the discussion on what an accountability system should look like and why. In considering policies on educational accountability, agents of accountability can use the framework to deliberate and be introspective about such policies. Where do states (or schools or districts) wish to be on each of these policy dimensions and what are the anticipated consequences of those decisions? It is perhaps consistent with Macpherson's (1996) methodological orientation regarding the development of accountability policies:

Begin not with definitions and theory but with competitive definitions and theories... End not with a rational and functional theory of appropriate behaviors based on objective facts but with a practical set of policy options that best accommodate empirical, subjective, and normative data until an even more comprehensive and more coherent account of educative accountability develops. (p. 81)

At its political core, accountability is about ensuring public tax dollars are spent wisely and that schools are meeting expectations of performance. That said, the question of how the educational establishment goes about this pursuit still remains. In this paper I have argued for policies that promote a balance of internal and external forms of accountability on the grounds that such policies are more symbiotic, more democratic, and more likely to lead to school improvement. Moreover, policies that define and measure performance holistically, that provide formative feedback to schools, that de-emphasize standardization and promote individuality in students, and do so with reasonable expectations and supportive consequences, are the most likely to bring about real school improvement.

In some respects, it all boils down to the question regarding the purpose of schools. The goals of U.S. schools are grounded in sand—they are disparate, numerous, and unagreed upon—and most defy accurate measurement of their attainment. To wit, Donna Kerr reminds us that “it is necessary to recognize that policies with unachievable purposes are amenable only to formative evaluation” (1976, p. 146). Contemporary models of educational accountability, in some insidious sense, run counter to the ultimate desire for our children to achieve the unattainable.

Notes

¹ The purported alignment between state academic standards and state assessments has been questioned by Popham (2001) and Achieve, Inc. (*Standards and accountability: Strategies for*

maintaining momentum, 2001), among others.

² See, for example, Vermont's Frameworks for Standards and Learning Opportunities, available online at <http://www.state.vt.us/educ/pdf/framework.pdf>

³ The framework was first developed from an inductive analysis of the accountability and assessment schemes of ten states (see Cobb, 2002). Thorough inspection of formal state policies and procedures identified multiple criteria on which accountability systems can be understood, judged, and compared.

⁴ Here is an example where states have a choice to make, at least with respect to the NCLB provisions.

⁵ "School Accountability for Learning and Teaching," Rhode Island Department of Elementary and Secondary Education, available online at <http://www.ridoe.net/schoolimprove/salt/default.htm>.

⁶ I do not intend to use the term "narrow" in a pejorative sense.

⁷ For instance, it is not uncommon to find references to "interpersonal skills," "working with others," or "a passion for the arts" in the language of state learning standards.

⁸ Incidentally, this strategy is typically based on an expected pattern of linear growth that all schools—regardless of their baseline level of performance—are expected to achieve. Let's say 2012 is the deadline for all schools to achieve 100% proficiency. For a school with 50% of its students achieving up to standard in 2002, the expectation is for that percentage to increase by 5% a year for the next 10 years—irrespective of the quality of the students that will attend that school during the next decade. This allowance of incremental growth conveys a sense of fairness and patience, but, at the same time, also puts off what could be the inevitable—that is, the notion that at least some schools won't make the goal.

⁹ Anand Vaishnav, "Appeals process offered on MCAS," *Boston Globe*, January 23, 2002.

¹⁰ For example, fewer than half the 1,019 teachers surveyed in a nationally representative poll said they had "plenty" of access to curriculum guides or other instructional materials that align with state standards. A similar percentage said they had "plenty" of access to training in the use of state standards or assessments (Quality Counts 2001, p. 9.).

¹¹ Quality Counts 2001, p. 8.

¹² "Implementing the Colorado Basic Literacy Act," available online at http://www.cde.state.co.us/cdeassess/download/pdf/asimp_cbla.pdf.

¹³ See the American Educational Research Association's statement on high stakes testing, available online at: <http://www.aera.net>.

¹⁴ For example, see Jane Hannaway, "How and why money matters: An analysis of Alabama schools," *Holding schools accountable: Performance-based reform in education*, ed. Helen F. Ladd (Washington, DC: Brookings Institution, 1996). Also see Benjamin Scafidi, Catherine Freeman, and Stan DeJarnett, "Local flexibility within an accountability system," *Education Policy Analysis Archives*, vol. 9(44), available online at <http://epaa.asu.edu/epaa/v9n44.html>.

¹⁵ William A. Firestone and David Mayrowetz, "Rethinking 'high stakes': Lessons from the United States and England and Wales," *Teachers College Record*, 102(4), August 2000, pp. 724-749.

¹⁶ See, for example, David Tyack and Larry Cuban's *Tinkering toward utopia: A century of school reform*, Cambridge, MA: Harvard University Press, 1995.

¹⁷ Action Planning Guide, Vermont Department of Education, available online at <http://>

www.state.vt.us/educ/actplan/APguide1.htm.

¹⁸ Lorraine Keeney, "Using data for school improvement," Report on the Second Practitioners' Conference for Annenberg Challenge sites, Annenberg Institute for School Reform, May 1998, p.40.

References

- Abelmann, C. H., Elmore, R. F., Even, J., Kenyon, S., & Marshall, J. (1999). *When accountability knocks, will anyone answer?* (RR-42): University of Pennsylvania, Consortium for Policy Research in Education.
- Bogdan, R. C., & Biklen, S. K. (2003). *Qualitative Research for Education* (4th ed.). Boston: Allyn & Bacon.
- Brooks, J., & Brooks, M. (1993). *In search of understanding: The case for constructivist classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Cobb, C. D. (2002). Performance-based accountability systems for public education. Online available at: <http://www.unh.edu/nhcpps/caseyreport.pdf>.
- Clarke, T. (2001). *Graduate exit exams*. National Governors Association. Retrieved February 12, 2001, from the World Wide Web: www.nga.org/center
- Delaware Department of Education (2004). Retrieved on October 7, 2004, from <http://www.doe.state.de.us/AAB/ACCTFAQ.pdf>.
- Elmore, R. F. (1990). *Restructuring schools: The next generation of educational reform*. San Francisco: Jossey-Bass.
- Elmore, R. F., Abelmann, C. H., & Fuhrman, S. H. (1996). The new accountability in state education reform: From process to performance. In H. F. Ladd (Ed.), *Holding schools accountable: Performance based reform in education* (pp. 65-98). Washington, DC: The Brookings Institution.
- Gardner, H. (2000). *The disciplined mind: Beyond facts and standardized tests, the K-12 education that every child deserves*. New York: Penguin Books.
- Gardner, H. (1991). *The unschooled mind: How children think & how school should teach*. New York: Basic Books.
- Goertz, M. E., & Duffy, M. C. (2001). *Assessment and accountability across the 50 states* (RB-33): Consortium for Policy Research in Education.
- Goertz, M. E., & Duffy, M. C. (2003). Mapping the landscape of high-stakes testing and accountability programs. *Theory into Practice*, 42(1), 4-11.
- Hall, P. M., & McGinty, J. W. (1997). Policy as the transformation of intentions: Producing program from statute. *The Sociological Quarterly*, 38(3), 439-467.
- Haas, E., Cobb, C. D., Rallis, S. F., & Wilson, G. (unpublished manuscript). *Standing in the way of high-quality learning: How No Child Left Behind misdirects our resources*.
- Kerr, D. H. (1976). *Educational policy: Analysis, structure, and justification*. McKay.
- Kirst, M. (1990). *Accountability: Implications for state and local policymakers*. Washington, DC: U. S. Department of Education, Office of Educational Research and Improvement, Information Services.
- Kuchapski. (1998). Conceptualizing accountability: A liberal framework. In R. J. Macpherson (Ed.), *The politics of accountability* (pp. 185-196). Thousand Oaks, CA: Politics of Education Association.
- Lipsky, M. (1980). *Street-level bureaucracy*. New York: Russell Sage Foundation.

- Macpherson, R. J. S. (1996). Educative accountability policy research: Methodology and epistemology. *Educational Administration Quarterly*, 32(1), 80-106.
- Maine Department of Education (2004). Retrieved October 7, 2004, from <http://www.state.me.us/education/lsalt/LAD/>.
- McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- McNeil, L. M. (2000). Creating new inequalities: Contradictions of reform. *Phi Delta Kappan*, 81(10), 728-734.
- New England Compact (2004). Retrieved on October 7, 2004, from <http://www.necomcompact.org/ea/gle.htm>.
- Newmann, F. M., King, M. B., & Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review*, 67, 41-74.
- Popham, W. J. (2001, April). *Scrutinizing high-stakes test items from an instructional perspective*. Paper presented at the National Council of Measurement in Education, Seattle, WA.
- Rallis, S. F., & MacMullen, M. M. (2000). Inquiry-minded schools: Opening doors for accountability. *Phi Delta Kappan*, 81(10), 766-773.
- Scafidi, B., Freeman, C., & DeJarnett, S. (2001). Local flexibility within an accountability system. *Education Policy Analysis Archives*, 9(44).
- Scriven, M. (1997). Truth and objectivity in evaluation. In W. R. Shadish (Ed.), *Evaluation for the 21st century: A handbook* (pp. 477-500). Thousand Oaks, CA: Sage Publications.
- Seeking stability for stands-based education. (2001, January 11, 2001). *Education Week*, XX, 8-9.
- Standards and accountability: Strategies for maintaining momentum* (2001). Achieve, Inc. Retrieved October 22, 2001, from the World Wide Web: www.achieve.org
- Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2002). *Survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost*. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
- Wiske, M. (1997). *Teaching for understanding: Linking research with practice*. San Francisco: Jossey-Bass.